MILLIMAN REPORT

Prevalence of rare disease in a commercial population using ICD-10 diagnosis codes

August 2022

Melanie K. Kuester, PharmD, BCPS
Jessica M. Naber, FSA, MAAA
Conner E. Smith, MS

Milliman

Table of Contents

Executive summary

In the United States (U.S.), the number and percentage of drugs approved to treat rare diseases has increased over the last five years. The objective of this research is to determine the prevalence of rare diseases in a U.S. commercial population using the Orphanet repository and claims data.

Orphanet is a prominent repository for worldwide rare disease registry data. This analysis relied upon Orphanet's data sets and mapping of ORPHAcodes (unique codes corresponding to rare diseases and subtypes) to International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10) diagnosis codes. A retrospective analysis of Orphanet's ICD-10 diagnosis codes was performed using the calendar year (CY) 2019 IBM MarketScan Commercial Claims Database to estimate the prevalence of rare disease in a U.S. commercial population.

Approximately 6,800 rare diseases were documented in Orphanet. Of those with a documented worldwide prevalence rate, 96% of rare diseases were ultra-rare and occur in less than 1 per 1,000,000 individuals. In the claims analysis, 19% of the population was observed to have at least one ICD-10 code associated with a rare disease. Of these, only 16% had a diagnosis associated with a disease-specific code; all others were nonspecific or broadly defined codes.

Identifying rare diseases in claims data using the currently available ICD-10 diagnosis codes is imprecise because most ICD-10 codes associated with rare diseases are too broadly defined. Potential solutions may include adaptation of ICD-11 rare disease diagnosis codes into ICD-10, incorporation of ORPHAcodes as a coding descriptor, or implementing a comprehensive rare disease research repository for the U.S. population.

Regenerative and targeted therapies increase the need to be able to accurately identify rare diseases. Increased specificity in ICD-10 codes for more rare diseases could enable more accurate identification.

Objective

The objective of this study was to determine the prevalence of rare diseases observed in a retrospective U.S. commercial claims database using International Classification of Diseases Tenth Revision (ICD-10) diagnosis codes from Orphanet's rare disease repository. The analysis intended to gain insight into the usefulness and accuracy of using ICD-10 diagnosis codes to identify patients with rare diseases.

Background and significance

The unlocking of the human genome provided the first real opportunity to understand, differentiate, and identify many rare diseases. At present, over 6,800 rare diseases have been identified in humans.[1] In the United States, a rare disease affects fewer than 200,000 individuals.[1,2] Although any single rare disease is by definition rare, it is estimated that 25 million to 30 million Americans have a rare disease.[1] The severity of rare diseases ranges from those that have significant impact on morbidity and mortality, such as spinal muscular atrophy type 1, to those that do not alter one's life span, such as retinitis pigmentosa.[3] For many rare diseases, scientists have now progressed beyond understanding and identifying the cause in our genetic makeup to developing therapies that specifically target some of these conditions.

Until the Orphan Drug Act (ODA) was passed in 1983, very few therapies were developed to treat rare diseases due to the small eligible populations.[4] The ODA, and its subsequent amendments, provided incentives for manufacturers to develop therapies for rare diseases.[5] Yet these products often come with a high cost. Since 2017, the median treatment cost, using wholesale acquisition cost, for up to one year for a product with an orphan indication was over $300,000 in 2021.[6] The price often reflects the small patient size and complexity for developing the therapy.

Notable among drugs treating rare diseases are gene and cell therapies, which first began to enter the U.S. market in 2017.[7] These therapies are administered over a short timeframe and use a patient's own genes or cells to treat or cure the disease.[8] This is opposed to most traditional chronic drugs that manage disease symptoms and require long-term or lifetime usage.[8] At the time of writing, the most expensive of these is Rethymic, a cell therapy made from human thymus tissue. When implanted, it will provide lacking thymus functioning in individuals with congenital athymia. It has a list price of $2.73 million.[9]

Because many drugs indicated for rare diseases are expensive, health insurers and other payers want to identify individuals with rare diseases in their covered populations to better understand current costs and estimate the budget impact of a new treatment, if one is nearing approval. However, identifying these members is not straightforward, as the United States lacks a central repository of rare disease data.

In the United States, the Genetic and Rare Diseases Information Center (GARD) hosts information about rare diseases, but other organizations such as the National Organization for Rare Diseases (NORD) also provide similar information.[10,11] Neither the government-led GARD nor the private sector NORD include a database for U.S. prevalence based on patient registries. The National Center for Advancing Translational Science (NCATS) launched an effort to create registries called the Rare Diseases Registry Program (RaDaR).[12] However, the initiative encourages individual organizations to create their own registries without mention of a single national database, to which advocacy organizations would report summarized demographic data such as prevalence, gender mix, age mix, etc.[12] Thus, any available data regarding U.S. prevalence must be gathered from disparate locations, with many estimates being difficult to find or unpublished.

Because a rare disease repository specific to the United States does not exist, this study utilized a central repository of rare disease research from Orphanet, an organization established in France by the French National Institute for Health and Medical Research (INSERM) in 1997.[13] In 2000, it switched from solely a focus on France to a European enterprise. Its purpose is to gather knowledge on rare diseases in order to improve the diagnosis, care, and treatment of patients with rare diseases.[13] Although the definition of a rare disease in Europe is different in absolute terms from the definition in the United States, both refer to the same case rate of 5 or fewer cases per 10,000 individuals.[14] Today, Orphanet is a prominent repository for data related to rare diseases across the world and includes prevalence information. Contributions to Orphanet's data primarily come from its 41 member countries, and while some of the information is obtained from U.S. sources, the United States is not a member country.[13]

Materials and methods

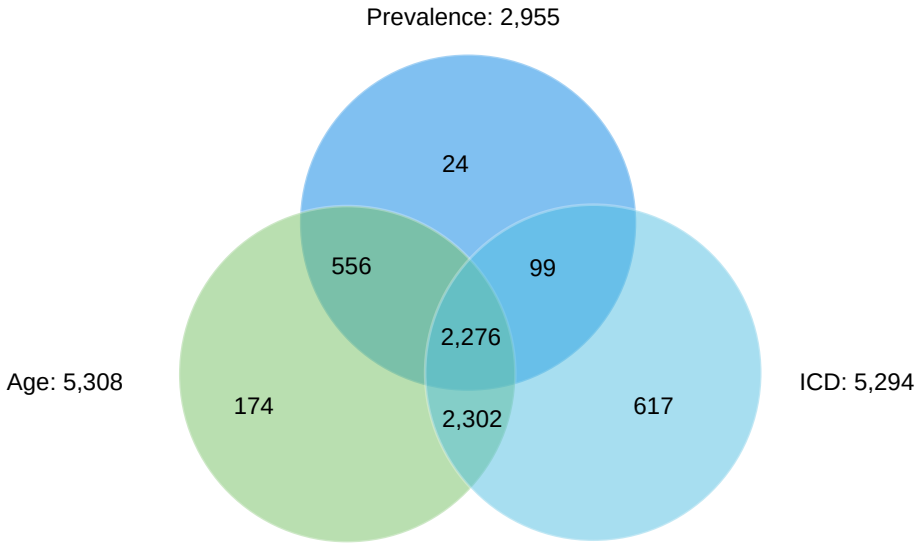
PHASE 1: ANALYSIS OF ORPHANET RARE DISEASE DATA

Orphanet creates a unique and stable identifier for each rare disease. These identifiers are called ORPHAcodes.[13,15] The data provided by Orphanet is called Orphadata and includes a crosswalk from ORPHAcodes to ICD-10 diagnosis codes.

Information is managed daily and changes can be seen in the Orphanet searchable database.[13,16] Monthly data is made available for download.[17] This research utilized the data sets "Rare Diseases and Alignment With Terminologies and Databases," "Epidemiology of Rare Diseases," and "Natural History of Rare Diseases" for the list of rare diseases and corresponding ICD-10 codes, prevalence, and age of onset information, respectively.[18,19,20] For this research, Orphadata from January 2021 was retrieved, organized, and analyzed using R.[21]

ORPHAcodes can correspond to medical conditions (e.g., hemophilia A, hemophilia B, acquired hemophilia), groups of conditions (e.g., hemophilia), and subtypes of conditions (e.g., mild hemophilia A, symptomatic form of hemophilia B in female carriers). For this analysis, only ORPHAcodes at the medical condition level (i.e., "Disorder" in Orphadata) were retained as they are "the main typological level … used to establish the total number of rare diseases that exist."[15] Worldwide point-prevalence estimates were retained for this study, as was the earliest recorded age of onset. Rare diseases without worldwide point-prevalence, age, or ICD-10 diagnosis codes were not included in this study. Figure 1 displays the availability of prevalence, age of onset, and ICD-10 information for rare diseases included in Orphadata.

FIGURE 1: COUNT OF ORPHADATA RARE DISEASES BY AVAILABLE METRICS, N=6,823*



*775 rare diseases were missing all of these metrics.

The mapping of ORPHAcodes associated with ICD-10 codes was used to facilitate the analysis of rare diseases in claims data (Phase 2). When multiple ICD-10 codes were assigned to a single ORPHAcode, the most specific ICD-10 code(s) was retained. To better assess the specificity of the available ICD-10 codes, the authors assigned each ICD-10 code to one of three categories: specific to a disease, associated with a broader body system or organ, or nonspecific. Orphanet changed the ICD-10 code for one ORPHAcode shortly after the data extraction date. Mature onset diabetes of the young, ORPHAcode 552, moved from E119 to E13. This change was incorporated in this analysis.

PHASE 2: ANALYSIS OF RARE DISEASES IN U.S. CLAIMS DATA

IBM MarketScan Commercial Claims and Encounters Database; 2019 4th Quarter Release was used to identify individuals who may have had a rare diseases diagnosis during CY2019 in the U.S. commercial population. This data set consists of medical and pharmacy administrative claims data from contributing employers and health plans for employees and their families who are covered by private health insurance. The total enrolled population in the Q4 2019 data set is approximately 25.5 million members.

The rare disease cohort in this study includes members who had at least one observation of an ICD-10 code from the Phase 1 ICD-10 rare disease code list, excluding laboratory and radiology claims. The breakdown of members by age range (0-19, 20-39, and 40+ years) and prevalence of the top rare disease ICD-10 codes were observed for both the rare disease cohort and the enrolled population.

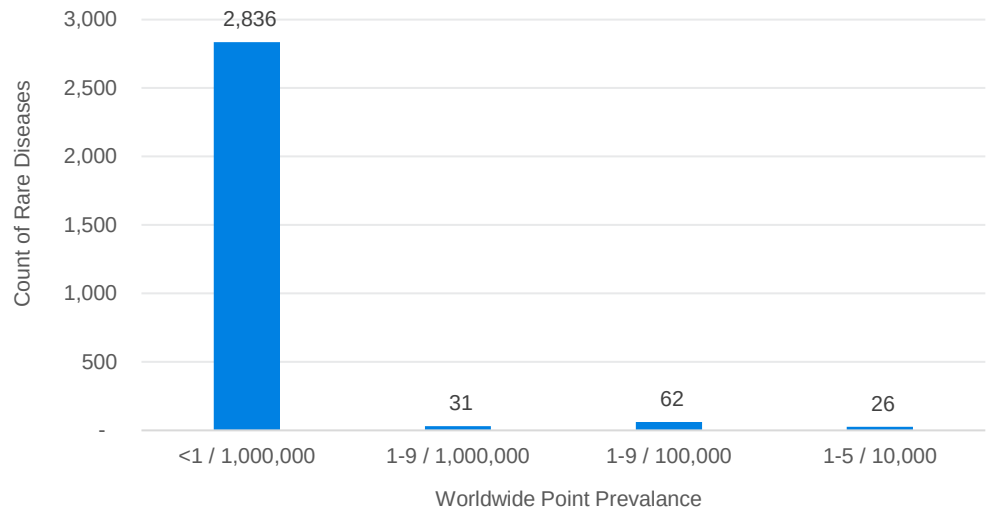
An individual was only counted once for a single rare disease. However, members with multiple ICD-10 codes (i.e., multiple rare disease diagnoses) were included in the counts of each associated ICD-10 code.

Results

PHASE 1: ANALYSIS OF ORPHANET RARE DISEASE DATA

Of the approximately 6,800 rare diseases included in Orphadata, 2,955 were found to have a documented worldwide prevalence rate. Figure 2 displays the distribution of diseases by prevalence rate. Nearly 96% of rare diseases with a documented worldwide prevalence rate are categorized as having a prevalence of less than 1 per 1,000,000. Only 1% of rare diseases have a prevalence rate of 1 to 5 per 10,000. When examining case count rather than prevalence rate, 25% of rare diseases (1,723 of 6,823) were documented as affecting fewer than a dozen people worldwide.

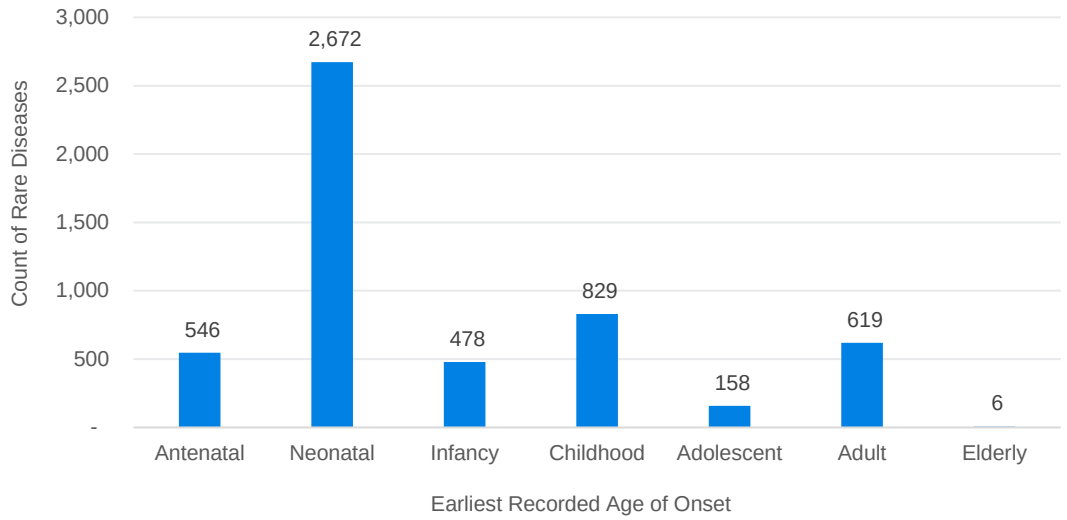
FIGURE 2: DISTRIBUTION OF RARE DISEASES BY WORLDWIDE PREVALENCE RATE*



* Prevalence available for 2,955 of 6,823 rare diseases.

Rare diseases can affect people of any age, but the majority have the earliest recorded onset during the first year of life (3,696 rare diseases out of 5,308 with a recorded onset). Figure 3 displays the number of rare diseases by the earliest recorded age of onset. Of those occurring in the first year, 75% have the earliest recorded onset occurring before or during infancy. The neonatal period is the single most frequent time period for a rare disease to be noticeable and potentially diagnosed (50%). Only 12% of rare diseases are categorized with the earliest age of onset as an adult.

FIGURE 3: DISTRIBUTION OF RARE DISEASES BY EARLIEST RECORDED AGE OF ONSET*



* Age of onset available for 5,308 of 6,823 rare diseases. Antenatal is before birth, neonatal is from birth to the fourth week of life, infancy is from the fourth week to the 23rd month, childhood is 2 to 11 years, adolescent is from 12 to 18 years, adult is 19 to 65 years, and elderly is 65 years and older.

An analysis of the ORPHAcode mapping to ICD-10 revealed that the assigned diagnosis codes are broadly defined for most rare diseases. Approximately 5,300 of the 6,800 rare diseases had one or more ICD-10 diagnosis codes assigned to it. Of the rare diseases with an assigned ICD-10 diagnosis code, 10% (n = 550) had an ICD-10 code that was specific to a disease, such as hereditary factor VIII deficiency (e.g., hemophilia A). Nearly 80% (n = 4,168) of rare diseases had ICD-10 codes that categorized the disease to a general body system or organ (e.g., Rabson-Mendenhall syndrome maps to "Other specified diabetes mellitus"), and the remaining conditions had an ICD-10 code that was nonspecific (n = 585), i.e., "Other specified congenital malformation syndromes, not elsewhere classified." The last two categories are broadly defined. For example, there are multiple conditions—including non-rare conditions—that a provider may code as "Other specified diabetes mellitus." This suggests it may be difficult to accurately identify the rare diseases that were mapped to the broadly defined ICD-10 codes.

PHASE 2: ANALYSIS OF RARE DISEASES IN U.S. CLAIMS DATA

In the commercial claims data, the total enrolled population (n = 25,486,307) had 27% of members aged 40 years or older, 33% between ages 20 and 39, and 40% under 20 years old. The split between gender was nearly equal, with 49% of enrollees being male and 51% female.

Using the resulting ICD-10 diagnosis codes and mapping from the Phase 1 analysis, 19% (n = 4,810,424) of individuals in the commercial claims data were observed to have at least one ICD-10 code linked to a rare disease and were included in the rare disease cohort for this research. The majority of the rare disease cohort were adults, with 64% of members aged 40 years or older, 22% between ages 20 and 39, and 14% under 20 years old. There was a higher proportion of females, 56%, versus males, 44%.

For the individuals in the rare disease cohort, 67% (n = 3,231,637) were observed to have only one rare disease diagnosis code, 22% (n = 1,052,615) had two rare disease diagnosis codes, 7% (n = 354,206) had three rare disease diagnosis codes, and the remainder had four or more rare disease diagnosis codes. Thus, one-third of individuals in the rare disease cohort had a diagnosis for more than one distinct rare disease during CY2019.

It was observed that 16% (n = 790,176) of the individuals in the rare disease cohort were observed with a specific diagnosis code, 92% (n = 4,448,621) were observed with a broadly defined diagnosis code, and fewer than 0.5% (n = 16,332) were observed with a nonspecific code. As previously noted, members with more than one distinct rare disease code could be counted in multiple categories but are only counted once within each category.

The top 10 rare disease ICD-10 chapters observed within the commercial population by age band, as defined by the World Health Organization (WHO), is displayed in Figure 4. More than a quarter of members (27.5%, n = 185,329) aged 0 to 19 years who had a rare disease were observed to have a diagnosis code categorized within the "Congenital malformation, deformation, and chromosomal abnormality" ICD-10 chapter. The top chapters found for individuals with a rare disease in the 20-to-39 age group were "Neoplasms" (30.4%, n = 314,442) and "Diseases of the skin and subcutaneous tissue" (25.6%, n = 264,527). The top chapters found for individuals with a rare disease in the 40-and-older age group were "Endocrine, nutritional, and metabolic diseases" (40.8%, n = 1,264,556) and "Neoplasms" (29.2%, n = 906,155).

FIGURE 4: TOP ICD-10 CHAPTERS BY AGE BAND FOR COMMERCIAL MEMBERS AMONG THOSE WITH RARE DISEASE DIAGNOSIS CODE*

ICD-10 CHAPTER (TOP 10 BY AGE GROUP)	% WITH ICD-10 CODE IN CHAPTER AGES 0-19	RANK IN AGES 0-19	% WITH ICD-10 CODE IN CHAPTER AGES 20-39	RANK IN AGES 20-39	% WITH ICD-10 CODE IN CHAPTER AGES 40+	RANK IN AGES 40+
	n = 673,082		n = 1,034,383		n = 3,100,051	
Congenital malformations, deformations and chromosomal abnormalities	27.5%	1	6.0%	7	2.9%	10
Neoplasms	14.8%	3	30.4%	1	29.2%	2
Endocrine, nutritional and metabolic diseases	7.2%	5	22.0%	3	40.8%	1
Diseases of the skin and subcutaneous tissue	19.3%	2	25.6%	2	25.5%	3
Certain infectious and parasitic diseases	9.6%	4	NA	NA	NA	NA
Diseases of the circulatory system	4.0%	10	6.1%	6	9.0%	4
Diseases of the genitourinary system	NA	NA	6.9%	4	NA	NA
Diseases of the musculoskeletal system and connective tissue	NA	NA	6.6%	5	7.6%	5
Diseases of the ear and mastoid process	5.9%	6	NA	NA	4.5%	7
Mental, Behavioral and Neurodevelopmental disorders	4.8%	7	NA	NA	NA	NA
Diseases of the digestive system	4.8%	8	4.8%	9	3.1%	9
Diseases of the nervous system	4.3%	9	5.7%	8	4.6%	6
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	NA	NA	4.3%	10	4.2%	8

* Members with multiple rare disease diagnosis codes may be included in multiple disease categories. Note that 2,908 members had a rare disease diagnosis but no identified age information.

Note: "NA" is displayed if the ICD-10 Chapter was not in the top 10 for the given age group.

The 25 most frequent rare disease diagnosis codes are displayed in Figure 5. Figure 5 includes the frequency of occurrence within the rare disease cohort and within the total enrolled commercial population. Approximately 21.2% (n = 1,017,436) of individuals in the rare disease cohort were observed to have a "Mixed hyperlipidemia" diagnosis code but this represents only 4.0% of the total commercial population. "Melanocytic nevi of the trunk" (13.4%, n = 645,332) and "Other melanin hyperpigmentation" (10.2%, n = 488,445) were the second and third most common rare disease diagnosis codes within the rare disease cohort, respectively.

FIGURE 5: TOP 25 RARE DISEASE ICD-10 DIAGNOSIS CODES FOR COMMERCIAL MEMBERS *

ICD-10 DIAGNOSIS CODE	DESCRIPTION	% WITH ICD-10 CODE AMONG THOSE WITH RARE DISEASE DIAGNOSIS	% WITH ICD-10 CODE AMONG TOTAL POPULATION**
		n = 4,810,424	n = 25,486,307
E782	Mixed hyperlipidemia	21.2%	4.0%
D225	Melanocytic nevi of trunk	13.4%	2.5%
L814	Other melanin hyperpigmentation	10.2%	1.9%
D229	Melanocytic nevi, unspecified	5.6%	1.1%
M722	Plantar fascial fibromatosis	5.4%	1.0%
L578	Oth skin changes due to chr expsr to nonionizing radiation	4.4%	0.8%
E291	Testicular hypofunction	4.0%	0.8%
E538	Deficiency of other specified B group vitamins	3.7%	0.7%
H903	Sensorineural hearing loss, bilateral	3.0%	0.6%
L538	Other specified erythematous conditions	2.4%	0.5%
L905	Scar conditions and fibrosis of skin	2.3%	0.4%
L918	Other hypertrophic disorders of the skin	2.3%	0.4%
E042	Nontoxic multinodular goiter	2.1%	0.4%
H538	Other visual disturbances	1.8%	0.3%
D239	Other benign neoplasm of skin, unspecified	1.7%	0.3%
D224	Melanocytic nevi of scalp and neck	1.5%	0.3%
I471	Supraventricular tachycardia	1.5%	0.3%
L728	Other follicular cysts of the skin and subcutaneous tissue	1.2%	0.2%
D508	Other iron deficiency anemias	1.2%	0.2%
B372	Candidiasis of skin and nail	1.2%	0.2%
C61	Malignant neoplasm of prostate	1.1%	0.2%
D126	Benign neoplasm of colon, unspecified	0.9%	0.2%
G35	Multiple sclerosis	0.9%	0.2%
K130	Diseases of lips	0.9%	0.2%
L603	Nail dystrophy	0.9%	0.2%

* Members with multiple rare disease diagnosis codes may be included in multiple diagnosis categories.

**Rare diseases have a prevalence rate of ≤ 0.06%.

For the ICD-10 codes identified as disease-specific, the top 10 ICD-10 diagnosis codes observed within each age band can be seen in in Figure 6. “Plagiocephaly” (Q673) was the most frequent disease-specific diagnosis for individuals aged 0 to 19, and “Plantar fascial fibromatosis” (M722) was the most frequent disease-specific diagnosis for individuals aged 20 to 39 and 40 and older.

FIGURE 6: DISEASE-SPECIFIC TOP 10 ICD-10 CODES IN COMMERCIAL BY AGE BAND*

ICD-10 CODE	DIAGNOSIS	% WITH ICD-10 CODE AMONG THOSE WITH RARE DISEASE DIAGNOSIS			% WITH ICD-10 CODE AMONG TOTAL POPULATION		
		AGES 0-19	AGES 20-39	AGES 40+	AGES 0-19	AGES 20-39	AGES 40+
		n = 673,082	n = 1,034,383	n = 3,100,051	n = 10,219,285	n = 8,413,028	n = 6,853,994
Q673	Plagiocephaly	3.30%	NA	NA	0.22%	NA	NA
Q211	Atrial septal defect	3.04%	0.44%	0.33%	0.20%	0.05%	0.15%
M722	Plantar fascial fibromatosis	1.80%	5.14%	6.21%	0.12%	0.63%	2.81%
P220	Respiratory distress syndrome of newborn	1.12%	NA	NA	0.07%	NA	NA
L813	Cafe au lait spots	0.94%	NA	NA	0.06%	NA	NA
Q909	Down syndrome, unspecified	0.90%	NA	NA	0.06%	NA	NA
H905	Unspecified sensorineural hearing loss	0.69%	0.48%	0.80%	0.05%	0.06%	0.36%
Q315	Congenital laryngomalacia	0.68%	NA	NA	0.04%	NA	NA
Q256	Stenosis of pulmonary artery	0.49%	NA	NA	0.03%	NA	NA
Q100	Congenital ptosis	0.42%	NA	NA	0.03%	NA	NA
N803	Endometriosis of pelvic peritoneum	NA	1.06%	NA	NA	0.13%	NA
M329	Systemic lupus erythematosus, unspecified	NA	0.91%	0.72%	NA	0.11%	0.32%
N800	Endometriosis of uterus	NA	0.86%	0.61%	NA	0.11%	0.27%
G932	Benign intracranial hypertension	NA	0.60%	NA	NA	0.07%	NA
D473	Essential (hemorrhagic) thrombocythemia	NA	0.59%	0.51%	NA	0.07%	0.23%
N801	Endometriosis of ovary	NA	0.51%	NA	NA	0.06%	NA
G501	Atypical facial pain	NA	0.50%	0.33%	NA	0.06%	0.15%
D869	Sarcoidosis, unspecified	NA	NA	0.45%	NA	NA	0.20%
G500	Trigeminal neuralgia	NA	NA	0.36%	NA	NA	0.16%
C541	Malignant neoplasm of endometrium	NA	NA	0.34%	NA	NA	0.15%

* Members with multiple rare disease diagnosis codes may be included in multiple diagnosis categories. Note that 2,908 members had a rare disease diagnosis but no identified age information.

Note: “NA” is displayed if the ICD-10 diagnosis code was not in the top 10 for the given age group.

Discussion

Health insurers and payers primarily rely on claims data and historical experience to estimate or project future costs and utilization. Using ICD-10 diagnosis codes is one of the most common ways to identify members in claims data who have been diagnosed with a particular condition, but there are limitations. This analysis demonstrated that it is difficult to identify members diagnosed with many rare diseases solely by using the current ICD-10 diagnosis codes because most rare disease codes are too broadly defined.

It was observed that a rare disease could be present in 19% of the sample commercial population when using only ICD-10 codes for identification. Comparatively, the published prevalence estimate in the United States for someone to have any rare disease is approximately one in 10.[22] In absolute terms, the results suggest that the total number of Americans with any rare disease would be greater than one in 10.

A key contributor to this is that 92% of individuals in the rare disease cohort were observed to be diagnosed with a broadly defined ICD-10 code. This means individuals with rare diseases are likely coded with the same ICD-10 diagnosis as members who have other, potentially non-rare conditions (but have a similar relationship to an overarching organ or body system). For patients with rare diseases that do not have a specific ICD-10 code available, it will be challenging for payers to identify these patients using the using the currently available ICD-10 codes.

Identification of individuals who have a rare disease becomes particularly relevant to health insurers and payers when there is a potential treatment option in development that may address an unmet need but comes with a high cost, such as a gene or cell therapy. Accurately identifying the treatment-eligible population is important for planning, pricing, and budgeting purposes. Using the current ICD-10 codes to identify rare diseases in claims data can help narrow the population, but they will not give an accurate view for most rare diseases because nearly 80% of rare diseases had ICD-10 codes that categorized the disease to a general body system or organ. Additionally, it was observed in the Phase 1 analysis that more than 20% of the 6,800 rare diseases identified in the Orphadata do not have any assigned ICD-10 code. This analysis underscores the need for more rare diseases to be issued specific diagnosis codes for both research and categorization purposes.

Incorporating additional criteria such as the use of certain treatments or procedures may aid in identifying members who are potentially eligible. However, 95% of rare diseases do not have a treatment approved by the U.S. Food and Drug Administration (FDA), which can make using drug or medical service utilization ineffective to adequately identify appropriate members.[23] Thus, multiple data sources, including but not limited to claims records, electronic health records (EHR), and underwriting reports may be necessary to identify members who are potentially eligible for a new treatment option. For this reason, the process to identify the treatment-eligible population can become onerous. For example, a member with hemophilia A can easily be identified in claims data because there is a specific diagnosis code and patients can be treated with hemophilia-specific therapies. However, the hemophilia A gene therapies in development have additional clinical criteria that reduce the treatable population even further, such as a minimum level of disease severity, a required number of bleeds or level of current treatment, and no prior history of inhibitors.[24,25] Eligibility criteria for other gene and cell therapies in development include adequate liver function, failure on one or more alternate treatments, and meeting certain thresholds from lab results.

Below are some potential solutions that could address these challenges and make it easier for payers and other healthcare stakeholders to better understand or identify rare diseases in the United States:

- ICD-11 incorporates specific codes for approximately 5,400 rare diseases.[26] The United States could adapt these additional codes for use in the ICD-10 coding system in order to provide more granularity for payers. ICD-11 was adopted by the World Health Organization in February 2022 and uses a different nomenclature.[27] Adaptation would allow the current coding processes and systems to be retained but provide necessary granularity into these conditions.

- The United States could adopt Orphanet's rare disease ORPHAcodes as qualifiers to nonspecific ICD-10 codes. The ORPHAcodes would provide further specificity about rare diseases, including severity. This would enable greater accuracy in identifying patients with severe forms of a rare disease, who are most likely to be indicated for a new therapy.
- There is currently no comprehensive rare disease repository for the U.S. population. Creating a repository specific to the United States or joining Orphanet as a member country could make it easier for researchers to gather information about rare diseases and their prevalence in the United States. These resources could be used by a payer to estimate the expected number of cases within its population and gain an understanding of demographic or clinical criteria that would refine the number of potentially treatment-eligible patients for a new therapy.

Changing the current system will come with additional administrative costs to healthcare stakeholders. However, if any of the above solutions were to be adopted by the United States and incorporated into the current coding of rare diseases—such as in claims data—it would benefit patients, providers, payers, and manufacturers of innovative therapies for rare diseases. Individuals with rare diseases could be more clearly identified, which enables better research and potential development of more treatments for rare diseases.

- **Patients** could have more streamlined paths to accessing treatments and could be identified easily for connections to care managers or advocacy groups.
- **Providers** could have a more efficient and effective way to connect with other providers treating patients with similar conditions or to communicate diagnoses during transitions of care. More granularity in coding could potentially align reimbursement to the level of severity of the patient's condition.
- **Payers** could better identify patients with rare diseases in their memberships to enable a greater understanding of their financial risks in upcoming plan years, such as for budgeting or rate setting.
- **Manufacturers** of rare disease therapies could better estimate the expected patient population size associated with rare diseases when developing new therapies. It could also facilitate enrollment into clinical trials for therapies being developed to treat rare diseases.

Limitations

This analysis relied on data from Orphanet and used the mapping from ORPHAcodes to ICD-10 codes. If there are material defects in the mapping, it is possible that they would be uncovered by a detailed, systematic review to search for code assignments that are questionable or materially inconsistent. Such a review was beyond the scope of this research.

It was also assumed that the diagnosis codes assigned would be the same code applied by a U.S. medical provider. Similarly, the data represented in the commercial claims database is reflective of ICD-10 codes assigned by providers at individuals' appointments and may not accurately reflect the best ICD-10 code to describe an individual's medical condition. Given that rare diseases may not be consistently coded, only one observation of a diagnosis code was required in the year, excluding lab or radiology claims.

Using claims data can be advantageous for analyses because of the large, diverse sample sizes, longitudinal nature of the data, and more comprehensive medical histories. Limitations when using claims data include reliance on accurate coding by providers and the limited information reported to payers; not all clinical information is captured on claims forms. This analysis was performed using claims data from the commercially insured population and may not be representative of the experience of any one specific commercial payer. It was assumed to be representative of the U.S. commercially insured population. The authors also anticipate the results are not likely to reflect the experience in other insurance channels, such as Medicare or Medicaid.

Conclusion

This study found that approximately 19% of the commercially insured population was observed to potentially have a rare disease when ICD-10 diagnosis codes were used for identification. Published estimates indicate one in 10 Americans has a rare disease; thus, using ICD-10 codes alone may considerably overstate the actual number of rare disease cases due to the lack of specificity for rare diseases in the current code set. Adaptation of ICD-11 rare disease diagnosis codes into the ICD-10 coding structure, adoption of Orphanet's ORPHAcodes as qualifiers to ICD-10 codes, and/or the creation of a rare disease repository specific to the United States are potential solutions to address the challenge of accurately identifying rare diseases within a claims data set.

Until there are condition-specific diagnosis codes available for every rare disease, it will be difficult for U.S. payers and other healthcare stakeholders to preemptively identify treatment-eligible members for a rare disease therapy in claims data using ICD-10 codes. Health insurers and researchers will need to assess each rare disease individually, be well-informed about current and future therapies, and use additional resources beyond diagnosis codes to best estimate the number of potential treatment-eligible members.

CONTACT

Melanie K. Kuester
melanie.kuester@milliman.com

Jessica M. Naber
jessica.naber@milliman.com

Conner E. Smith
conner.smith@milliman.com

REFERENCES

1. National Human Genome Research Institute. Rare Diseases FAQ. Retrieved July 26, 2022, from https://www.genome.gov/FAQ/Rare-Diseases.
2. GARD Genetic and Rare Diseases Information Center. About GARD: FAQs about rare diseases. Retrieved July 26, 2022, from https://rarediseases.info.nih.gov/diseases/pages/31/faqs-about-rare-diseases.
3. NORD® (National Organization for Rare Disorders). Rare Disease Database: Werdnig-Hoffmann Disease. Retrieved July 26, 2022, from https://rarediseases.org/for-patients-and-families/information-resources/rare-disease-information/.
4. Huron J. (March 25, 2021). New Study Investigates the Number of Available Orphan Products, Generics and Biosimilars. NORD. Retrieved July 26, 2022, from https://rarediseases.org/new-study-investigates-the-number-of-available-orphan-products-generics-and-biosimilars/.
5. FDA (February 20, 2020). Rare Diseases at FDA. Retrieved July 26, 2022, from https://www.fda.gov/patients/rare-diseases-fda.
6. Kuester, M, & Fohl, Z. (January 2022). Milliman Medicaid pipeline and new drug report: Q1 2022.
7. Novartis (August 30, 2017). Novartis receives first ever FDA approval for a CAR-T cell therapy, Kymriah™ (CTL019), for children and young adults with B-cell ALL that is refractory or has relapsed at least twice. Retrieved July 26, 2022, from https://www.novartis.com/news/media-releases/novartis-receives-first-ever-fda-approval-car-t-cell-therapy-kymriahtm-ctl019-children-and-young-adults-b-cell-all-refractory-or-has-relapsed-least-twice.
8. American Society of Gene and Cell Therapy. Gene Therapy 101: Different Approaches. Retrieved July 26, 2022, from https://patienteducation.asgct.org/gene-therapy-101/different-approaches.
9. FDA (October 8, 2021). FDA Approves Innovative Treatment for Pediatric Patients with Congenital Athymia. Retrieved August 12, 2022, from https://www.fda.gov/news-events/press-announcements/fda-approves-innovative-treatment-pediatric-patients-congenital-athymia.
10. National Center for Advancing Translational Sciences. Search for a Disease. Genetic and Rare Diseases (GARD) Information Center. Retrieved July 26, 2022, from https://rarediseases.info.nih.gov/.
11. National Organization for Rare Disorders (NORD). Retrieved July 26, 2022, from https://rarediseases.org/.
12. National Center for Advancing Translational Sciences. Rare Diseases Registry Program (RaDaR). Retrieved July 26, 2022, from https://ncats.nih.gov/radar.
13. Orphanet. About Orphanet. Retrieved July 26, 2022, from https://www.orpha.net/consor/cgi-bin/Education_AboutOrphanet.php?lng=EN.
14. Orphanet (September 2021). Orphadata Free Access Products Description. Retrieved July 26, 2022, from http://www.orphadata.org/cgi-bin/img/PDF/OrphadataFreeAccessProductsDescription.pdf.
15. Orphanet (March 2020). Procedural document: Orphanet nomenclature and classification of rare diseases, Version 02. Retrieved July 26, 2022, from https://www.orpha.net/orphacom/cahiers/docs/GB/eproc_disease_inventory_R1_Nom_Dis_EP_04.pdf.
16. Orphanet. The portal for rare diseases and orphan drugs. Retrieved July 26, 2022, from https://www.orpha.net/consor/cgi-bin/index.php.
17. Orphadata. Welcome to Orphadata: Access to Aggregated Data From Orphanet. Retrieved July 26, 2022, from http://www.orphadata.org/cgi-bin/index.php#page-top.
18. Orphanet. Rare Diseases and Alignment With Terminologies and Databases. Retrieved January 8, 2021, from http://www.orphadata.org/cgi-bin/rare_free.html.
19. Orphanet. Epidemiology of Rare Diseases. Retrieved January 8, 2021, from http://www.orphadata.org/cgi-bin/epidemio.html.
20. Orphanet. Natural History of Rare Diseases. Retrieved January 8, 2021, from http://www.orphadata.org/cgi-bin/epidemio.html.
21. R. The R Project for Statistical Computing. Retrieved July 26, 2022, from https://www.R-project.org/.
22. About GARD: FAQs, op cit.
23. Rare Disease Day. Rare Disease Day: Frequently Asked Questions. Retrieved July 26, 2022, from https://rarediseases.org/wp-content/uploads/2019/01/RDD-FAQ-2019.pdf.
24. ClinicalTrials.gov (November 16, 2021). Single-arm study to evaluate the efficacy and safety of valoctocogene roxaparvovec in hemophilia A patients (BMN 270-301). Identifier: NCT03370913. Retrieved July 26, 2022, from https://clinicaltrials.gov/ct2/show/NCT03370913.
25. ClinicalTrials.gov (July 8, 2021). Safety, tolerability, and efficacy study of valoctocogene roxaparvovec in hemophilia A with active or prior inhibitors. Identifier: NCT04684940. Retrieved July 26, 2022, from https://clinicaltrials.gov/ct2/show/NCT04684940.
26. Ayme, S., Bellet, B., & Rath, A. (March 26, 2015). Rare diseases in ICD-11: Making rare diseases visible in health information systems through appropriate coding. Orphanet Journal of Rare Diseases. Retrieved July 26, 2022, from https://ojrd.biomedcentral.com/articles/10.1186/s13023-015-0251-8.
27. World Health Organization. ICD-11 2022 Release. Retrieved July 26, 2022, from https://www.who.int/news/item/11-02-2022-icd-11-2022-release.